

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 15:49:44 ; Search time 1021.63 Seconds

(without alignments)
216.534 Million cell updates/sec

Title: US-09-593-914-1

Perfect score: 15

Sequence: 1 agcgggtctataga 15

Scoring table:

IDENTIFY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hqo_hum:*
20: em_hqo_inv:*
21: em_hqo_fod:*
22: em_hq_hum1:*
23: em_hq_hum2:*
24: em_hq_hum3:*
25: em_hq_hum4:*
26: em_hq_hum5:*
27: em_hq_hum6:*
28: em_hq_hum7:*
29: em_hq_hum8:*
30: em_hq_inv1:*
31: em_hq_inv2:*
32: em_hq_other:*
33: em_hq_fod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_vl3:*
61: gb_hq2:*
62: gb_hq3:*
63: gb_hq4:*
64: gb_hq5:*
65: gb_hq6:*
66: gb_hq7:*
67: gb_hq8:*
68: gb_hq9:*
69: gb_hq10:*
70: gb_hq11:*
71: gb_hq12:*
72: gb_hq13:*
73: gb_hq14:*
74: gb_hq15:*
75: gb_hq16:*
76: gb_hq17:*
77: gb_hq18:*
78: gb_hq19:*
79: gb_hq20:*
80: gb_hq21:*
81: gb_hq22:*
82: gb_hq23:*
83: gb_hq24:*
84: gb_hq25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	9 AX057545	AX057545 Sequence
2	15	100.0	1787	13 BB18BNA	X83815 B.bruxellien
3	15	100.0	1799	13 BB18BNA	X83814 B.bruxellien
4	15	100.0	1799	13 BB18BNA	X83805 D.bruxellien
5	15	100.0	198038	65 AC018980	AC018980 Homo sapi
6	15	100.0	209213	66 AC020644	AC020644 Homo sapi
7	14	93.3	39547	64 AC016460	AC016460 Homo sapi
8	14	93.3	45300	12 AC007288	AC007288 Arabidops
9	14	93.3	55739	13 AP000389	AP000389 Arabidops
10	14	93.3	81414	12 AB024037	AB024037 Arabidops
11	14	93.3	110157	12 AF058825	AF058825 Arabidops


```

FEATURES             97089000      Location/Qualifiers
source               1..1799
                    /organism="Brettanomyces bruxellensis"
                    /strain="MCC 362 non type strain"
                    /db_xref="taxon:37784"
                    /rname="1..1799"
                    /product="18S ribosomal RNA"
                    /partial
BASE COUNT          476 a 358 c 487 g 475 t      3 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DH 13; Length 1799;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agcgagctctataga 15
    |||
Db 1065 ACCCGCTCTATTAGA 1051

RESULT 4
B18SRKNA/c
LOCUS      B18SRKNA      1799 bp      DNA
DEFINITION D.bruxellensis gene for 18S ribosomal RNA.
ACCESSION  X58052
VERSION     X58052.1 GI:2485
KEYWORDS   18S ribosomal RNA.
SOURCE      Dekkera bruxellensis.
ORGANISM    Dekkera bruxellensis.
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Dekkera.
REFERENCE   1 (bases 1 to 1799)
AUTHORS     De Wachter R.
TITLE       Direct Submission
JOURNAL     Submitted (22-FEB-1991) R. De Wachter, Universiteit Antwerpen
            (UIA), Universiteitsplein 1, 2610 Wilrijk, Belgium
REFERENCE   2 (bases 1 to 1799)
AUTHORS     Hendriks L., Goris A., Van de Peer Y., Neels J.M., Vancanneyl M.,
            Kersters K., Berty J.F., Hennebert G. and De Wachter R.
TITLE       The evolution of ascomycetes and ascomycete-like yeasts as inferred
            from small ribosomal subunit RNA sequences
JOURNAL     Syst. Appl. Microbiol. 15, 98-104 (1992)
FEATURES             Location/Qualifiers
source               1..1799
                    /organism="Dekkera bruxellensis"
                    /strain="MCC 27700"
                    /db_xref="taxon:5007"
                    /rname="1..1799"
                    /product="18S rRNA"
                    /gene="18S rRNA"
                    /product="18S ribosomal RNA"
                    /gene="18S rRNA"
BASE COUNT          479 a 359 c 483 g 478 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DH 13; Length 1799;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agcgagctctataga 15
    |||
Db 1066 ACCCGCTCTATTAGA 1052

RESULT 5
AC018980/c
LOCUS      AC018980      198038 bp      DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-476F14, WORKING DRAFT
SOURCE     9 unordered pieces.

```

```

ACCESSION      AC018980
VERSION        AC018980.5 GI:9887658
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Human.
ORGANISM       Homo sapiens
               Hominidae; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 198038)
AUTHORS        Smith,D.R.
TITLE          Genome Therapeutics Corporation Sequencing Center: Human genome
               Sequence Data
               Unpublished
               2 (bases 1 to 198038)
REFERENCE      2 (bases 1 to 198038)
AUTHORS        Smith,D.R.
TITLE          Direct Submission
JOURNAL        Submitted (25-DEC-1999) genome Therapeutics Corporation, 100 Beaver
               Street, Waltham, MA 02453, USA
               On Aug 24, 2000 this sequence version replaced gi:828944.
COMMENT        Genome Center
               Center: genome Therapeutics Corporation
               Center code: GTC
               Web site: http://www.genomecorp.com/
               Contact: gtc-sequencer@genomecorp.com
               Project Information
               Center project name: hg081
               Summary Statistics
               Sequencing vector: N/A
               Chemistry: Dye-terminator Big Dye 100% of reads
               Assembly program: Phrap; version 990315
               Consensus quality: 181125 bases at least Q40
               Consensus quality: 191464 bases at least Q30
               Consensus quality: 193045 bases at least Q20
               Insert size: 197298; sum-of-contrigs
               Quality coverage: 5.1x in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3331: contrig of 3231 bp in length
3332
3332: gap of unknown length
3332
11374: contrig of 8043 bp in length
11375
11475: gap of unknown length
11475
25209: contrig of 13735 bp in length
25210
45310: gap of unknown length
45310
31680: contrig of 6376 bp in length
31680
42741: gap of unknown length
42741
42741: contrig of 10956 bp in length
42742
42842: gap of unknown length
42842
59298: contrig of 16457 bp in length
59299
59398: gap of unknown length
59399
89881: contrig of 30483 bp in length
89882
89981: gap of unknown length
89982
122341: contrig of 32360 bp in length
122342
122441: gap of unknown length
122442
198038: contrig of 75597 bp in length.
198038
Location/Qualifiers
1..198038
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-476F14"
/clone_id="RP11-11"
1..3231
/note="assembly_name:Contig23"
3332..11374
/note="assembly_name:Contig24"
11475..25209
/note="assembly_name:Contig25"

```


*	115394	125420:	contig of 10027 bp in length
*	125421	126520:	gap of unknown length
*	126522	134678:	contig of 8158 bp in length
*	134679	134778:	gap of unknown length
*	134779	141834:	contig of 7056 bp in length
*	141835	141934:	gap of unknown length
*	141935	150537:	contig of 8603 bp in length
*	150538	150637:	gap of unknown length
*	150638	165416:	contig of 15679 bp in length
*	165417	165416:	gap of unknown length
*	165417	183214:	contig of 16798 bp in length
*	183215	183314:	gap of unknown length
*	183315	209213:	contig of 25899 bp in length

FEATURES
source

```
/organism="Homo sapiens"  
/db_xref="Locus:9606"
```

BASE COUNT						
61464	a	42813	c	42287	q	58391
						t
						4258 others
ORIGIN						

Query Match	100.08;	Score 15;	HH 66;	Length 209213;
Host Local Similarity	100.08;	Pred. No. 52;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

QY      1  agcgcgtctataga 15
          |||||
IDb  16198  AGCGGCTCTATTAGA 16184

```

RESULT	7
AC016460	
LOCUS	39547 bp DNA
DEFINITION	Homo sapiens clone R11-10B1, LOW-PASS
ACCESSION	AC016460
VERSION	AC016460.2 GI:9123815
KEYWORDS	HTG; HTGS-PHASED.
SOURCE	human.
	13-JUL-2000
	SAMPLING.

ORGANISM	REFERENCE
Homo sapiens	Lukavtsova; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Plitheria; Primates; Carnivora; Hominoidea; Homo.
1 (bases 1 to 39547)	
Dierren, B., Linton, L., Nussbaum, C., and Lande, E.	
Homo sapiens, clone RPL1-10E1	
Unpublished	
2 (bases 1 to 39547)	
Biren, B., Linton, L., Nussbaum, C., Lande, E., Allen, N., Anderson, P.	

TITLE	Direct Submission
JOURNAL	Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genomes Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Jul 13, 2000 this sequence version replaced g1:6479045.

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR

```

Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L3050
Center clone name: 10_E_1

```

* NOTE: this record contains 44 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful to
* identify clones that may be generic and allow
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	810	909:	gap of	809	Contig of 809 bp in length
*	910	1691:	Contig of 782 bp in length		
*	1692	1791:	gap of	100 bp	
*	1792	2585:	Contig of 794 bp in length		
*	2586	2685:	gap of	100 bp	
*	2686	3491:	Contig of 806 bp in length		
*	3492	3591:	gap of	100 bp	
*	3592	4403:	Contig of 812 bp in length		
*	4404	4504:	gap of	100 bp	
*	4504	5316:	Contig of 813 bp in length		
*	5317	5416:	gap of	100 bp	
*	5417	6217:	Contig of 801 bp in length		
*	6218	6317:	gap of	100 bp	
*	6318	7105:	Contig of 788 bp in length		
*	7106	7205:	gap of	100 bp	
*	7206	8011:	Contig of 806 bp in length		
*	8012	8111:	gap of	100 bp	
*	8112	8900:	Contig of 789 bp in length		
*	8901	9000:	gap of	100 bp	
*	9001	9802:	Contig of 802 bp in length		
*	9803	9902:	gap of	100 bp	
*	9903	10675:	Contig of 773 bp in length		
*	10676	10775:	gap of	100 bp	
*	10776	11594:	Contig of 819 bp in length		
*	11595	11694:	gap of	100 bp	
*	11695	12495:	Contig of 801 bp in length		
*	12496	12595:	gap of	100 bp	
*	12596	13396:	Contig of 803 bp in length		
*	13399	13498:	gap of	100 bp	
*	13499	14322:	Contig of 824 bp in length		
*	14323	14422:	gap of	100 bp	
*	14423	15248:	Contig of 815 bp in length		
*	15249	15348:	gap of	100 bp	
*	15349	16141:	Contig of 804 bp in length		
*	16142	16241:	gap of	100 bp	
*	16242	17056:	Contig of 795 bp in length		
*	17057	17136:	gap of	100 bp	
*	17137	17953:	Contig of 817 bp in length		
*	17954	18053:	gap of	100 bp	
*	18054	18824:	Contig of 771 bp in length		
*	18825	18924:	gap of	100 bp	
*	18925	19731:	Contig of 807 bp in length		
*	19732	19831:	gap of	100 bp	
*	19832	20620:	Contig of 789 bp in length		
*	20621	20720:	gap of	100 bp	
*	20721	21511:	Contig of 791 bp in length		
*	21512	21611:	gap of	100 bp	
*	21612	22397:	Contig of 786 bp in length		
*	22398	22497:	gap of	100 bp	
*	22498	22308:	Contig of 811 bp in length		
*	23309	23408:	gap of	100 bp	
*	23409	24213:	Contig of 811 bp in length		
*	24220	24319:	gap of	100 bp	
*	24320	25111:	Contig of 794 bp in length		
*	25111	25231:	gap of	100 bp	

```

* 25214 26025: contig of 812 bp in length
* 26026 26125: gap of 100 bp
* 26126 26917: contig of 792 bp in length
* 26918 27017: gap of 100 bp
* 27018 27819: contig of 802 bp in length
* 27820 27919: gap of 100 bp
* 27920 28727: contig of 808 bp in length
* 28728 28827: gap of 100 bp
* 28828 29610: contig of 783 bp in length
* 29611 29710: gap of 100 bp
* 29711 30523: contig of 813 bp in length
* 30524 30623: gap of 100 bp
* 30624 31419: contig of 796 bp in length
* 31420 31519: gap of 100 bp
* 31520 32324: contig of 805 bp in length
* 32325 32424: gap of 100 bp
* 32425 33213: contig of 789 bp in length
* 33214 33313: gap of 100 bp
* 33314 34121: contig of 808 bp in length
* 34122 34221: gap of 100 bp
* 34222 35025: contig of 804 bp in length
* 35026 35125: gap of 100 bp
* 35126 35947: contig of 822 bp in length
* 35948 36047: gap of 100 bp
* 36048 36848: contig of 801 bp in length
* 36849 36948: gap of 100 bp
* 36949 37741: contig of 793 bp in length
* 37742 37841: gap of 100 bp
* 37842 38625: contig of 784 bp in length
* 38626 38725: gap of 100 bp
* 38726 39547: contig of 822 bp in length.

```

FEATURES

```

source
1. 39547
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="RP11-10E1"
   /clone_lib="RP11-10E1"
BASE COUNT 10768 a 7134 c 7055 g 10151 t 4439 others
ORIGIN

```

```

Query Match      93.3%  Score 14:  DE 64:  Length 39547;
Best Local Similarity 100.0%:  Pred. No. 2,3e-02;
Matches 14:  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Oy 1 agcgagctctatag 14
      |||||
Db 10427 AGCGGCTCTATAG 10440

```

```

RKSUJF 8
AC007288 45300 bp JNA PIN 05-AUG-2000
DEFINITION Arabidopsis thaliana chromosome II section 69 of 255 of the
complete sequence. Sequence from clones F10C8.
AC007288 AF002093
VERSION
AC007288.12 GI:5598747
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

```

REFERENCE

AUTHORS

Lin, X., Kaul, S., Kounsley, S.D., Shea, J.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldblyum, T.V., Buell, C.R., Kelchum, R.A., Lee, J.J., Ronning, C.M., Koo, H., Motil, K.S., Cronin, L.A., Shen, M., Vanaken, S.E., Umayam, L., Talton, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Freuss, D., Nerman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.

TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis

```

JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PubMed 10617197
REFERENCE 2 (bases 1 to 45300)
AUTHORS Lin, X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

```

COMMENT

On Dec 17, 1999 this sequence version replaced gi:4850411. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GAT1 (<http://artbur.epm.oral.gov/pub/karat/>), GeneLinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://www.stanford.edu/GENSCAN.html>), and N-PlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-se (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CGH/Mash/ABI consortium for sequencing BAC clones FEP23, F516, F1745, and F13L16, the ESA group for sequencing clone F1304, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tachata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Ixina Kou, Hani Kizlak, Michael E. Heavey, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES

```

source
1. 45300
   /organism="Arabidopsis thaliana"
   /submitter="TIGR"
   /db_xref="taxon:3702"
   /chromosome="II"
   complement(1..>45300)
   /note="Sequence from clone F10C8"
   /rpt_family="F10C8"
   /rpt_family="F10C8"
   /gene="At2g12650"
   /note="F10C8.11"
   /pseudo
   /rpt_region
   4389..4765
   /note="DNA sequence related to non-LTR Yall family of
   retrotransposons"
   9531..9641
   /rpt_family="F10C8"
   /rpt_family="F10C8"
   /gene="At2g12660"
   /note="F10C8.10"
   /pseudo
   12278..15124
   /gene

```

/gene="At2g12670"
 /note="p10c8.9"
 /pseudo
 repeat_region
 12647..12802
 /note="DNA sequence related to non-LTR Tail family of
 retrotransposons"
 repeat_region
 complement(19134..19261)
 /rpt_family="(TA)n"
 gene
 21297..25308
 /gene="At2g12680"
 /note="p10c8.8"
 /pseudo
 repeat_region
 complement(22159..22261)
 /rpt_family="(CAAA)n"
 repeat_region
 22681..22732
 /rpt_family="(GAA)n"
 repeat_region
 complement(26355..26475)
 /rpt_family="(TAAA)n"
 gene
 complement(26838..27372)
 /gene="At2g12690"
 /note="p10c8.7"
 /pseudo
 repeat_region
 complement(27694..35519)
 /note="this region contains 747 bp tandem repeat units,
 like the one in bp 23172..23914. There is a 20 kb gap
 between 27227 and 27286. The entire region (27 kb) thus
 contains about 37 copies of the repeat units."
 complement(31405..31464)
 /note="gap of about 20 kb. This gap contains pure tandem
 repeats of 747 bp. We are currently trying different
 strategies to try to close this gap."
 misc_feature
 repeat_region
 complement(36774..36993,37088..37323)
 /gene="At2g12700"
 complement(36774..36993,37088..37323)
 /note="p10c8.6: similar to GH:AD23022"
 CDS
 complement(join(36774..36993,37088..37323))
 /gene="At2g12700"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="A031080.1"
 /db_xref="GI:4850413"
 /translation="MNPIDGLPLKRVYAVLTPADPRADPLNPLGIGTWVIXITWTW
 KFNHQLAVTICDILQNRHHNMTPTKMPVYKASTICMHTWDSRHITLIVMOPNIVA
 KTLRDNRAKRAKRAAGDGRSMKMTTIVNDLVDMTEFGPDVSGMS"
 37987..39001
 /gene="At2g12710"
 /note="p10c8.5"
 /pseudo
 mRNA
 complement(join(39615..40076,40213..41085,41290..42414))
 /gene="At2g12720"
 complement(39615..42414)
 /gene="At2g12720"
 /note="p10c8.4: similar to MUKA transposase of maize
 Mutator transposon"
 CDS
 complement(join(39615..40076,40213..41085,41290..42414))
 /gene="At2g12720"
 /codon_start=1
 /product="Mutator-like transposase"
 /protein_id="A031079.1"
 /db_xref="GI:4850412"
 /translation="MAGPATVIVVSGSWKRRQREYVNDRCRCRVQLEAKAVYDSL
 LKMYVDQYQNGSTHQRSLSYKSKTKLKNPFTPTPTVTSNDRLQCLYLTKTEQL
 RLCEVFTAKDSDISRKDGSRGSKRPLESDSRRESRFEPQMGEGVGYDQDEA
 GDERGDGDSGVDADEKESDEEDDESRPFDDSDGASDEDEPAVGEPRKN
 DEDSPKSPSKRIQJMEMTOSAGSLTDLKLSTLTLAVCOQYRSFELYELKLLAI
 RDGEFDFTSNKTVSYKQWDRCLMRACRQGNPNFYVYVYNSRSTGCVARRSG
 RSROATPVYGVLYRDYLDGPDYKPKSICITITKPKRKASTSYSTLRRARELT
 IGTDSSEELPSTLYIKRANPGVYARLQDDEGRNMFYVYGASIGFHTMRVY
 VYDTLHGSKGLTLALQDGNFOIFPLAFVAVDENDSRMLFTLQAKVIDAT
 DUALISDHSKISGAEVYPLAAGICTYLLKYNLILYFKRRKDLPLVKKARCYRL
 NPTNAENLEEDPLILHAYLQAGVEMWARRPGDRYVNLMTNNAESMNAISLQAK
 NLPVIRMEIAIROMMTWMAFERRDASQHTQLPVGEXKILQTRVTSNLLDVQIDA

repeat_region
 BASH COUNT 1414 a 7779 c 8978 g 14339 t 60 others
 ORIGIN

Query Match 93.3%; Score 14; DB 12; Length 45500;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 ggcggctcattaga 15
 |||||
 Db 40071 GCGGGTCAATTAGA 40084

RESULT 9
 LOCUS AP000389 55739 bp DNA PLN 27-DEC-2000
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MSJ3.
 ACCESSION AP000389 BA000014
 VERSION AP000389.1 GI:5672589
 KEYWORDS
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_jib:Mitsui P1
 clone:MSJ3.
 ORGANISM Arabidopsis thaliana
 Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II;
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS Kaneko,T., Kato,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
 TITLE 1 (sites)
 JOURNAL Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
 MEDLINE 2 (bases 1 to 55739)
 REFERENCE DNA Res. 7 (3), 217-221 (2000)
 AUTHORS Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1532-3, Yama,
 Kisarazu, Chiba 292-0812, Japan (E-mail:yakamukazusa.or.jp,
 Tel:81-438-52-3935, Fax:81-438-52-3934)

COMMENT

For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MSJ3
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graal
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Graal-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 McGene2 (S.M. Hebsgaard, et al., CSH, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/McGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremind.zool.lastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MVAL1 and the 3' clone is MOP15.

FEATURES

source
 Location/Qualifiers
 1..55739
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MSJ3"
 /clone_jib="Mitsui P1"

CDS	join(462, .735,1345, .1373) /note="gb AAC61810.1 gene_id:MSJ3.1 similar to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB01420.1" /db_xref="GI:11994245" /translation="MDPKAEISRLVLPDPLGEIYAKIGAKYADYINCLSKEL GASADEKVLKTLNLAFLVKPLSCSKLILMKKGLANNPDNHYTKANSRNER" complement(join(2420, .4168,4328, .5023,5138, .5459)) /note="contains similarity to Arabidopsis thaliana retrotransposon Athlia gene_id:MSJ3.2" /codon_start=1 /pseudo /evidence-not_experimental 6784, .8561 /note="gene_id:MSJ3.3 unknown protein" /codon_start=1 /pseudo /evidence-not_experimental 8666, .9013 /note="gb AA24530.1 gene_id:MSJ3.4 similar to unknown protein" /codon_start=1 /pseudo /evidence-not_experimental join(13427, .13652,13605, .13769,14564, .15219,15302, .15409, 15500, .15556,15681, .15725,15808, .16049,16127, .16235, 16688, .16859,16997, .17124,17227, .17382) /note="gb AAC97213.1 gene_id:MSJ3.5 strong similarity to unknown protein" /codon_start=1 /pseudo /evidence-not_experimental join(19136, .20065,20144, .20368,20467, .20550,20807, .20926, 21101, .21220,21809, .21955,22036, .22182,22263, .22353, 22432, .22550,22624, .22794,22894, .22941) /note="gb AA25853.1 gene_id:MSJ3.6 strong similarity to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB01421.1" /db_xref="GI:11994246" /translation="MSIKLIPPKRIPLVGSVNIITRVSRSSEKIVSVKALIP ENTLMDFLPVFEADLAFSSHIVHLLJRRILTNDELMEVADQPKSLKE PLITGLPMDASSSVNKRKKKKMMIKDNTLKSKKQVWIKREMTPLILTYVL EKRSKKMPEERLESLVGLILLARNPTMLPTENLKATDEKPKKYMAYIY SVLVHVKRIGNDYILRKQYGLPGVSGTOWALSSVOKIGKHGCRDSSGNPII LQMSSTSPSAKALIMADGKIVHPLSCGPRVYLQKPLFTDPRKRIKDIIDPVYI KQKRSKKDWCGMIOVBOGHOORIDHVAKAIITVHOSRRKRSPOKRRKRRG KKQDEEKKQEEGKEELEKVEYKRGDGTKEQEIPIKQDEMEGEEKQEEGEEK BEKVEYKGDGSTERQEIPIKQDEMEVEEEKQEEGEEKVEKEVEEHEE TEDEAVYIISDEDNGTAPTEKESQPKETETVEKRENEHDEHDETEQEA LSDDEDNTAPTEKESQPKETETVEKRENEHDEHDETEQEA PLTPVMAISQVMEIDIAVKKAKKKILRLDLSLPHILMPISKKQKYLIMMDERK YKAIKKATILIPDKILAAHRYVYIVQICKILVKKDRAKAKMDIVL" complement(join(23766, .24009,24128, .25699,25824, .25874, 26025, .26075,26157, .26453)) /note="gene_id:MSJ3.7" /codon_start=1 /pseudo /evidence-not_experimental /product="Mutator-like transposase" 33238, .33464 /note="contains similarity to En/Spm-like transposon protein gene_id:MSJ3.8"
CDS	/codon_start=1 /pseudo /evidence-not_experimental 39943, .40386 /note="gene_id:MSJ3.9 unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB01422.1" /db_xref="GI:11994247" /translation="MAEFHLPLNTEPDQTEDEFEADTLLIDDEYRLAYLDSR SHNGVTCGVDEILFQILVGNSSSCRRMILNAGEHLPVVEFTAFEMHKGLVCAICRE ELAAARLISPLPKCRHYHKKCIISNMLSNNTICPIORHAYELPNNC" 42049, .47107 /note="gb AAD15542.1 gene_id:MSJ3.10 strong similarity to unknown protein" /codon_start=1 /pseudo /evidence-not_experimental complement(join(47610, .47819,47908, .48078,48155, .48202)) /note="gb AAD34021.1 gene_id:MSJ3.11 similar to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB01423.1" /db_xref="GI:11994248" /translation="MAORSLFGESSDDEELDRMEIRMTRYGERRRDRDQVPIED CNKVVAVTSLDVPWTEKLPFSCPYEISIGDQCGCPKRWVTAICQEFDMIKETSE MKDILHANKKHYEQAKILPLMKKPELLEKKYHSLANKYI" join(49962, .50126,50196, .50501) /note="gb AAC69115.1 gene_id:MSJ3.12 similar to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB01424.1" /db_xref="GI:11994249" /translation="MSSNSSENSASTQSVSRNGRLPKSKGCHGIDVYITSTAKRPG REPERPTRKDDILFKWEDGTIEVVVDSWRLSVNDELSAAKSEVALELQAMV KHLKPKVMSRSPENQIMKMLKICQCLAPTIIVGLIMPKINPRLILGY" complement(join(51037, .51156,51256, .51486)) /note="contains similarity to Ku/Spm-like transposon protein gene_id:MSJ3.13" /codon_start=1 /evidence-not_experimental /protein_id="BAB01425.1" /db_xref="GI:11994250" /translation="MVFHILKQELKIVEVYLTAHPRDLPVNLIGSTWADKTYTF KNNDLAVAILGDLIKNFHIAKITVGMILVVKRMTIDSHTLLTWQEPNLVARTLY KNNKALKRRRRVY" join(52402, .52436,52529, .52793,52867, .52969,53069, .53312, 53312, .53312) /note="gene_id:MSJ3.14" /codon_start=1 /evidence-not_experimental /product="replication protein A1-like" /protein_id="BAB01426.1" /db_xref="GI:11994251" /translation="MCDYGLRDIINPCITCMTICVAVRRMYISPTDENSEILWVVE WGQILKALISRPFOYQDMLENNQKELIRFGVLENGCFVRSITKRYMELMNTVE TISDREVELLNRPEDYIIEPTATENLVDLQALVDGCLRNITEDGMDGDFRL TEIMDKYKRELKCEAGELAVDHRKPRSPNSIYINDIFALIMWKLIFRYHGLK HKRPNVAGISDVTVPDNIPEVQELKMYE"
CDS	BASE COUNT 18008 a 10125 c 10679 g 16927 t ORIGIN
CDS	Query Match 93.3% Score 14: DB 13: Length 55739; Best Local Similarity 100.0%; Prid. No. 2.3e+02; Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2 qcgcgtctataga 15
 Db 54327 CCGCGTCTATTAGA 54340

RESULT 10
 AB024037
 LOCUS AB024037 81414 bp DNA PLN 09-AUG-2000
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MSK10.
 ACCESSION AB024037
 VERSION AB024037.1 GI:4519196
 KEYWORDS

ORGANISM
 Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
 clone:MSK10.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids 11; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty p1 and TAC clones
 JOURNAL DNA Res 7 (1), 31-63 (2000)
 MEDLINE 20181125
 RERERENCE 2 (bases 1 to 81414)
 AUTHORS Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (24-FEB-1999) to the DDBJ/EMBL/GenBank databases.
 Yasutazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yada, Kasarazu, Chiba 292-0812, Japan
 E-mail:y.nakamura@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934
 Address for correspondence: kasuekazusa.or.jp
 For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/seq_graph.cgi?c=MSK10
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Gra1 (Informatics Group, Oak Ridge National Laboratory), <http://combio.ornl.gov/Gra1-1.3/>, GENSCAN (Chris Burge, MIT, <http://cbr-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremmlini.zool.tasstate.edu/cu-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/cody/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MG623 and the 3' clone is T13612.

FEATURES
 source
 1.81414
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MSK10"
 /clone_lib="Mitsui P1"
 complement(1..192)
 /note="CDS is reported in Acc# AB028613
 gene_id:MG623.7"
 /number=1
 /evidence="not_experimental"

exon
 complement(241..2797)
 /note="CDS is reported in Acc# AB028613
 gene_id:MG623.7"
 /number=1
 /evidence="not_experimental"

CDS
 /product="retroelement pol polypeptide-like"
 /note="emb|CAH77928.1
 gene_id:MSK10.2
 similar to unknown protein"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BAH97394.1"
 /db_xref="GI:8843868"
 /translation="MIRIALHACVLSLSGLFLPASKLEPTSONKYSTPTSDP
 GAILPHTSTSSNLSCHSMWLLDPLVSHATPDVTRPPTLPRTSTWSSSTITTH
 HHHSSTHSITSSOSTQIRTPDKLSTRKRGEDKVGSGDLHRTKPLSAHYSWG
 RAIRLLARLSEFYMYR"
 join(4695..5268,5361..5486,5645..5724,5835..5960,
 6033..6137,6227..6328,6399..6458,6542..6629,6876..7058)
 /note="emb|CAH1580.1
 gene_id:MSK10.3
 strong similarity to unknown protein"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 7427..8119
 /note="contains similarity to mutator-like transposase
 gene_id:MSK10.4"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 8890..9582
 /note="contains similarity to mutator-like transposase
 gene_id:MSK10.5"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BAH97395.1"
 /db_xref="GI:8843869"
 /translation="MPNPIHPDILAKKAKKNCWYLCMKSGRCHDLYRPD
 VKDANFEMCDPDDRYAMGHDNDSDHREPKKQDAGKRTTIEGLRVT
 HESDIGAILHLERLAPRNAAFTCALILYICAGNMGVYLRLGFSNHALESSEA
 KDICEVEDIDIKYGNILKYTAMGSPGCGVSTPKCAEMCYRSRLNINLCNOCY
 IMICAKRIQSIL"
 join(10112..10220,10432..10664)
 /note="gene_id:MSK10.6
 similar to unknown protein"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BAH97396.1"
 /db_xref="GI:8843870"
 /translation="MKPGSTYMYQKTSQSPVKRREYITDVGAIMRGIDLPNI
 SILVNTLPVKGPTLETLLNLYSSDINTLAIDEMTWSLQPESTRPYFELLVERG
 SLKLVAKKATVS"
 complement(11072..11536)
 /note="gb|AAH24957.1
 gene_id:MSK10.7
 strong similarity to unknown protein"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 join(17864..18351,18409..18718,18998..19184,19258..19718
 20145..20803,20889..20997,21270..21941)
 /note="gene_id:MSK10.8"
 /codon_start=1
 /evidence="not_experimental"
 /product="retroelement pol polypeptide-like"
 /protein_id="BAH97397.1"
 /db_xref="GI:8843871"
 /translation="MSEDAKREYPPRIYQEGDSNBEKGINHNITLGEPAIDESIGK
 ELVAELKTELRLAKLVDYSFLWSKRTVHYLLQRLKLEKMLCTVAAPKTRPGLN
 EFHHTIGLIDELPEKFEVPAADYKAFSEHLYVGEENLDELROGLVTCRAMPEK
 RRM1PESAKRYVDKAMKTYPMQRAAYELAVYSIKRLRPGLGTYTSGLVAYLOMA
 YSELITLTAERFGNAVADEILLLWNGNTRATTSVIAEDIKANGELRPTWPKA
 EISLPHFWPDLEDPALDNTLQPLHGYKLVINGYDYVTSKKKKLLKDSRRKKKK
 KKKKKVAAYSEDEDEKRSRDQSGDGLVDVAVSNLTPSNRFEVDTSTKEWSS
 RLVTGQLESSVEAKREGFGSTLNDVQIKQELKATADSKSSSYTRDMFLAKPOPO

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 123080)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project
Unpublished
3 (bases 1 to 123080)
Walterston, R.
Direct Submission
Submitted (30-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
4 (bases 1 to 123080)
Walterston, R.
Direct Submission
Submitted (11-MAY-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
5 (bases 1 to 123080)
Walterston, R.
Direct Submission
Submitted (17-JUN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
6 (bases 1 to 123080)
Walterston, R.
Direct Submission
Submitted (13-AUG-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: r.walterston@wustl.edu

COMMENT

MAPING: Clones were assigned to the YAC map by hybridization by M. Jodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

NOTES:

The 3' clone is T6L9. Actual start of this clone is at base position 1 of F10A2; actual end is at 123080 of F10A2.

FEATURES
source
Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1..123080
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1Y"
/map="unknown"
/clone="F10A2"

repeat_region
46..544
/note="Athila_retroelement"
repeat_region
343..544
/note="H12_Athila_180_bp_repeat"
repeat_region
614..675
/note="Athila_retroelement"
repeat_region
614..675
/note="H12_Athila_180_bp_repeat"
repeat_region
706..797
/note="H12_Athila_180_bp_repeat"
repeat_region
706..797
/note="Athila_retroelement"
repeat_region
756..884
/note="H12_Athila_180_bp_repeat"
repeat_region
756..884
/note="Athila_retroelement"
repeat_region
929..958
/note="H12_Athila_180_bp_repeat"
repeat_region
987..1094
/note="Athila_retroelement"
repeat_region
988..1094
/note="H12_Athila_180_bp_repeat"
repeat_region
1077..1135
/note="H12_Athila_180_bp_repeat"
repeat_region
1077..1135
/note="Athila_retroelement"
repeat_region
1138..1259
/note="H12_Athila_180_bp_repeat"
repeat_region
2209..2234
/note="Athila_retroelement"
repeat_region
2245..2306
/note="Athila_retroelement"
repeat_region
2766..2867
/note="Athila_retroelement"
repeat_region
2906..3015
/note="Athila_retroelement"
gene
3188..6268
/gene="F10A2.16"
CDS
3016[3188..3449,3535..3949,4127..4617,4968..5329,5488..5631,5691..6268]
/gene="F10A2.16"
/note="contains similarity to a family of Arabidopsis thaliana hypothetical proteins, which have similarity to retrotransposon Athila"
/codon_start=1
/evidence=not experimental
/protein_id="F10A2.16"
/db_xref="GI:4773919"
/translation="MAGQNPVQVQDPNDQPLNIGVGDAPRNHHQGGIVPPVQNNFEIKSLQSLIKIKRKEFLGEPEDPLDLSPRLCGTRKINDFTNSMDCQKVEFLAKFSNVRPRLRNITSGHNGKNNKNNPPGKAMRPFKSYTCQCHHCPRKASLSTYRGLIKLIMLIDPNSNGNFKNDVRCRQWELVNIAGSDCCYNIDYRSYKGTGSGEDKQSDIKALNKASIVNANVQKYVPRQDQSKPVOVMQGVPPQGFNGQIQDQNSPDPFIQOPALNADQEDIKOLIQQATRAMAFDKRIAEIHNKIAELINKIDMSYNDLNVKEQANSTKIKVENQFASSTAPRHPQALPGRAVONPMDYATANAITHQEDSGPTKRLGIMETKFCNATALLFANGSVAPHHGLIENLEPKIYQNEIPIDFVVDVDEESQDPLIGPFIASAGAVLDVRNKNINLEKIGIKKKFDIPEAGVONPFGIQSMVDDELTETSPKMKPOLMPEPKRKYQKSNQIDPRMETEALAKEDNPTETMPFNAQDQGCATPYLNKICQKSLNQKKTKTTLVYKKKLLAKKQMSHTPIQDEMOSQIOTLKIQLAKRTSLFSSSTSLTVRVLSDLKQAHDEGPMPSIHVNFLLFLKFLKCRFVLTLRSTV

3234..3522
/note="Athila_retroelement"
3486..4181
/note="Athila_retroelement"
4201..4399
/note="Athila_retroelement"
4395..4615
/note="Athila_retroelement"
4753..5249
/note="Athila_retroelement"
6124..6216
/note="Athila_retroelement"

```

gene
7103..9908
/ gene="P10A2.15"
join(7103..7358,7496..8580,9633..9908)
/ gene="P10A2.15"
/ note="contains similarity to a family of Arabidopsis
Challiana hypothetical proteins; see GB:A077408"
/codon_start=1
/evidence="not-experimental"
/protein_id="A029788.1"
/db_xref="GI:4773918"
/translation="MREYVARKRYVEAESEHHEAPFEETVSEKVELIGDDPTP
VRRNPKRKRREPIPSYIYQYIAELK+FGTSYPRKQTMQREKREKGIQYITPVK
VWYTDICHLNTIPGPSPGSLMODYQELKSLMKTIQCKPSPYKSSSTINPV
IYIHOCLANTIPKRYGVSSEELICMIIDSLVFIAREKIDKMKMDKADISLV
LDHLSTVEYAAITIRSGIRGSLGCVGLTPIIGAVMEGEPDVLPKPIIDITLG
KDELETPADRYALFKFNHLELPSILPLCENOTSXKTKNINMSPALNITIG
AREYSAVDQDEYFTQADYEDJEDJDIITYLEGTSSEQANIEEGDQGTDLAE
LHKIGVLEICNIIATKQKFKKMKTKSKMAVDLKDQSKPPSPPIIAT
SVLEHAGIIMSNHIMKSSSKFISTDHSTVSSGCHVPIPSDHSMAHLLPVVC
IHLHCOITPRHSHMKSSITITPPRIATR"
repeat_region
8823..8877
/ note="Alh1a_retroelement"
8823..8877
/ note="H12_Ath1a_180_bp_repeat"
8874..8953
/ note="Ath1a_retroelement"
8874..8953
/ note="H12_Ath1a_180_bp_repeat"
9100..9162
/ note="H12_Ath1a_180_bp_repeat"
9320..9373
/ note="Ath1a_retroelement"
9320..9373
/ note="H12_Ath1a_180_bp_repeat"
9404..9471
/ note="H12_Ath1a_180_bp_repeat"
9404..9471
/ note="Ath1a_retroelement"
9404..9471
/ note="H12_Ath1a_180_bp_repeat"
10333..10356
/ note="Ath1a_LTR_A"
10389..10439
/ note="106H"
10400..10435
/ note="Ath1a_LTR_A"
10442..10644
/ note="106B"
10468..10581
repeat_region
Query Match 93.3% Score 14: DB 14: Length 123080;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 gcaqqlctatlaag 15
|||||
db 62997 GCGGCTATTAGA 63010

RESULT 13
AC026473 170390 bp DNA PRI 27-SEP-2000
LOCUS Homo sapiens chromosome 16 clone RP11-5106, complete sequence.
DEFINITION AC026473
AC026473.6 GI:10312259
VERSION
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 170390)
AUTHORS
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 170390)

```

```

AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
3 (bases 1 to 170390)
AUTHORS
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Submitted (27-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT
On Sep 27, 2000 this sequence version replaced gi:9256555.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
SHOC-85708 G52652.
FEATURES
location/Qualifiers
1..170390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-5106"
BASP COUNT 55301 a 30773 c 31277 g 53039 t
ORIGIN
Query Match 93.3% Score 14: DB 88: Length 170390;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acaqqlctatlaag 14
|||||
db 43885 AGCGGCTATTAG 43872

RESULT 14
AC009169 175462 bp DNA PRI 26-NOV-2000
LOCUS Homo sapiens chromosome 16 clone RP11-90K18, complete sequence.
DEFINITION AC009169
AC009169.7 GI:11344989
VERSION
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 175462)
AUTHORS
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
3 (bases 1 to 175462)
AUTHORS
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT
On Nov 26, 2000 this sequence version replaced gi:9256159.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
FEATURES
location/Qualifiers
1..175462
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


misc_feature /note="assembly_fragment:01928
fragment_chain:2"
78437..84166
/note="assembly_fragment:01930
fragment_chain:3"
84267..96814
/note="assembly_fragment:01918
fragment_chain:3"
96915..118894
/note="assembly_fragment:01110
fragment_chain:3"
118995..121637
/note="assembly_fragment:01381
fragment_chain:4"
121738..129786
/note="assembly_fragment:01421
fragment_chain:4"
129887..133215
/note="assembly_fragment:00105"
133316..138221
/note="assembly_fragment:00370"
138322..154648
/note="assembly_fragment:01029"
154749..158946
/note="assembly_fragment:01170"
159047..166727
/note="assembly_fragment:00541
fragment_chain:5"
166828..173339
/note="assembly_fragment:00224
fragment_chain:5"
173440..175742
/note="assembly_fragment:00840
fragment_chain:5"
clone_end:SP6
vector_side:right"
BASE COUNT 52688 a 33525 c 33326 g 54000 t 2203 others
ORIGIN

Query Match 93.3%; Score 14; DB 81; Length 175742;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 qcqggtctatlaqa 15
|||||
Db 31936 GCGGCTPATTAGA 31923

Search completed: May 1, 2001, 17:45:29
Job time: 6945 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 16:41:49 ; Search time 113.81 Seconds
(without alignments)
76.942 Million cell updates/sec

Title: US-09-593-914-1

Sequence: 15
1 agcgcgtctataga 15

Scoring table: IDENTITY_MUC
Gapop 10.0, Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database:

N.Geneseq_0401:*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SID56/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.4	89.3	1071	21	Arabidopsis thaliana
2	13.4	89.3	1085	21	Arabidopsis thaliana
3	13.4	89.3	3408	19	Dead Box X (DBX) g
4	13.4	89.3	5321	21	Human DBX1 coding
5	13.4	89.3	5322	19	Dead Box X (DBX) g
6	13.4	89.3	8957	16	HTLV-II virus NRA
7	13.4	86.7	360	11	Sequence encoding
8	13.4	86.7	396	11	Sequence encoding
9	13.4	86.7	843	11	Sequence encoding
10	13.4	86.7	1151	21	Aspergillus oryzae
11	13.4	86.7	1260	13	Sequence encoding

12	13	86.7	1782	17	T38241
13	12.4	82.7	533	20	X21072
14	12.4	82.7	1167	18	T79770
15	12.4	82.7	1167	18	T79791
16	12.4	82.7	1336	21	C69404
17	12.4	82.7	1407	21	C69404
18	12.4	82.7	1559	21	C69404
19	12.4	82.7	1815	21	C69404
20	12.4	82.7	2022	17	T16016
21	12.4	82.7	2127	17	T16016
22	12.4	82.7	2249	13	Q29710
23	12.4	82.7	2832	15	Q62186
24	12.4	82.7	4809	15	Q62186
25	12.4	82.7	7475	12	Q15312
26	12.4	82.7	11475	19	V59979
27	12.4	82.7	11478	18	T79784
28	12.4	82.7	11478	18	T79805
29	12.4	82.7	11478	18	V55069
30	12.4	82.7	13811	20	X20649
31	12.4	82.7	143068	21	F21105
32	12.4	82.7	143068	21	F21272
33	12.4	82.7	143068	21	A34983
34	12.4	82.7	143068	21	A35150
35	12.4	82.7	149412	21	A35151
36	12.4	82.7	152740	21	F21273
37	12.4	82.7	1830121	17	T42063
38	12	80.0	296	21	A79502
39	12	80.0	970	21	C47429
40	12	80.0	973	21	C33442
41	12	80.0	1167	19	Z96423
42	12	80.0	1167	19	V37401
43	12	80.0	1486	21	C51465
44	12	80.0	1488	21	C33887
45	12	80.0	1491	21	C76525

ALIGNMENTS

RESULT 1	
C46237	
ID	C46237 standard; DNA; 1071 BP.
AC	C46237;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49404.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
FE	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
XX	
PR	05-MAR-1999; 99US-0123180.
XX	
PR	09-MAR-1999; 99US-0123548.
XX	
PR	23-MAR-1999; 99US-0125788.
XX	
PR	25-MAR-1999; 99US-0126264.
XX	
PR	29-MAR-1999; 99US-0126785.
XX	
PR	01-APR-1999; 99US-0127462.
XX	
PR	06-APR-1999; 99US-0128234.
XX	
PR	08-APR-1999; 99US-0128714.
XX	
PR	16-APR-1999; 99US-0129845.
XX	
PR	19-APR-1999; 99US-0130077.
XX	
PR	21-APR-1999; 99US-0130449.

Mycoplasma 46-48 k
Polynucleotide seq
Potato starch bran
Potato starch bran
Human secreted pro
Human secreted pro
Mouse sphingosine
Mouse sphingosine
DNA encoding ubiq
Ub17 ubiquitin-lyt
PKC promoter from
Cyanobacteria repl
PA01 plasmid fragm
Plasmid pHP12 enc
Potato starch bran
Full length potat
Potato class B sta
Polynucleotide seq
Human low adenosi
Human low adenosi
Human adenosine re
Human adenosine re
Human adenosine re
Human low adenosi
Haemophilus influe
Pinus radiata cell
Arabidopsis thalia
Arabidopsis thalia
S. pneumoniae deri
Streptococcus pneu
Arabidopsis thalia
Arabidopsis thalia
Human ORF2080

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131444.
PR 30-APR-1999; 99US-0132040.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132481.
PR 05-MAY-1999; 99US-0132481.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132489.
PR 07-MAY-1999; 99US-0132861.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134211.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135352.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139111.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142890.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148561.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152463.
PR 10-SEP-1999; 99US-0153707.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 89.3%; Score 13.4; DB 21; Length 1071;
Best Local Similarity 93.3%; Pred. No. 41;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtctattaga 15
|||
Db 922 agtgggtctattaga 936

RESULT 2
ID C33698 standard: DNA; 1085 BP.
XX C33698;

XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3998.

XX Hybridisation assay: genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP103405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

```

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147320.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 89.3%; Score 13.4; DB 21; Length 1085;
Best Local Similarity 93.3%; Pred. No. 41;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtctataga 15
DB 919 agtgggtctataga 933

RESULT 3
ID V69632 standard; DNA: 3408 BP.
AC V69632;
XX V69632;
DT 02-FEB-1999 (first entry)
XX
DE Dead Box X (DBX) gene short transcript nucleic acid sequence.
XX
KW Non-recombining region; human; Y chromosome; X homologue; testis; DBX;
KW infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
XX
OS Homo sapiens.
XX
PN W09846747-A2.
XX
PD 22-OCT-1998.
XX
PF 10-APR-1998; 98WD-US07115.
XX
PR 11-APR-1997; 97US-0041877.
XX
PA (WHD) WHITEHEAD INST BIOMEDICAL RES.
PI Lahn BT, Page DC;
XX
DR WPI: 1998-568729/48.
XX
PT P-PSDB; W81502.
XX
PT Novel genes in the non-combining region of Y chromosome - useful to
PT diagnose if male infertility or reduced sperm count has a genetic
PT basis
XX
PS Disclosure: Fig 3A-B; 54pp; English.
XX
XX This represents the nucleotide sequence of the Dead Box X (DBX) gene
XX short transcript. The invention relates to genes occurring on the non-
XX recombining region of the human Y chromosome. The sequences fall into two
XX classes: (1) X-homologous DNA which are expressed in many organs, having
XX functional X homologues and (2) testis-specific DNA sequences. Y
XX chromosomal DNA from males with known conditions such as infertility and
XX reduced sperm count can be assessed using the invention to determine
XX whether the condition is associated with or caused by the occurrence of
XX the gene or gene alteration. Candidate inhibitors of the enzymatic
XX activity of the genes can be assessed using in vitro assays.
XX
SQ Sequence 3408 BP; 951 A; 734 C; 923 G; 800 T; 0 other;

```

Query Match 89.3%; Score 13.4; DB 19; Length 3408;
 Best Local Similarity 93.3%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
 ||||| |||||
 Db 728 agcgggtctattaga 742

RESULT 4
 A29207 ID A29207 standard; DNA; 5321 BP.
 XX
 AC A29207;
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Human DBX1 coding sequence.
 XX
 KW Topoisomerase III-alpha; interacting protein; TIII-a IP; RNA helicase;
 KW DBX1; nucleosome; supercoiling; chromosome segregation; recombination;
 KW stabilization; cell division; apoptosis; cell cycle regulation;
 KW cytotoxic; anti-tumour; DBX1; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT CDS 856..2844
 FT /*tag= a
 FT /standard_name= DBX1
 FT /product= RNA_helicase-like_protein
 XX
 PN M0200032768-A1.
 XX
 PD 08-JUN-2000.
 XX
 PD 29-NOV-1999; 99WO-FR02952.
 XX
 PR 30-NOV-1998; 98FR-0015081.
 XX
 PA (AVET) AVENTIS PHARMA SA.
 XX
 P1 Pournier A, Goulaouic H, Riou J;
 XX
 DR WP1: 2000-412316/35.
 DR P-PSDB; Y96483.
 XX
 PT New nucleic acid encoding ligand for topoisomerase III alpha, useful
 PT for inhibiting the enzyme and in drug screening, e.g. for potential
 PT anticancer agents
 XX
 PS Claim 6: Page 47-53; 68pp; French.
 XX
 CC This DNA encodes a DBX1, which shows homology with RNA helicases but the
 CC activity of a helicase has never been demonstrated and its function has
 CC not yet been identified. DBX1 possesses the 8 characteristic motifs of
 CC the helicases of the family "DEAD". In particular, it appears to be part
 CC of the sub-family represented by helicase PL10. The DBX1 gene is situated
 CC on the X chromosome and its homologue, which is situated on the Y
 CC chromosome has 91 percent identity with a novel protein coding sequence.
 CC The novel protein (see Y96482) is a human topoisomerase III-alpha
 CC interacting protein (TIII-a IP). TIII-a IP has structural features in
 CC common with RNA helicases (e.g. DBX1), which are involved in destruction
 CC of the nucleosome structure, supercoiling of DNA, segregation of newly
 CC replicated chromosomes, and recombination and stabilization of the
 CC genome. Agents that alter interaction between TIII-a IP and TIII-a
 CC therefore modulate cell division, replication, transcription,
 CC translation, splicing and DNA recombination or repair, so may slow cell
 CC growth, block the cell cycle or induce apoptosis. Antibodies and ligands
 CC of TIII-a IP are used to prevent, treat or alleviate diseases that
 CC involve abnormal regulation of the cell cycle, i.e. they are potential

CC anti-tumour (cytotoxic) agents.
 XX
 SQ Sequence 5321 BP; 1532 A; 1010 C; 1274 G; 1505 T; 0 other;

Query Match 89.3%; Score 13.4; DB 21; Length 5321;
 Best Local Similarity 93.3%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
 ||||| |||||
 Db 727 agcgggtctattaga 741

RESULT 5
 V69631 ID V69631 standard; DNA; 5322 BP.
 XX
 AC V69631;
 XX
 DT 02-FEB-1999 (first entry)
 XX
 DE Dead Box X (DBX) gene long transcript nucleic acid sequence.
 XX
 KW Non-recombining region; human; Y chromosome; X homologue; testis; DBX;
 KW infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
 XX
 OS Homo sapiens.
 XX
 PN M09846747-A2.
 XX
 PD 22-OCT-1998.
 XX
 PE 10-APR-1998; 98WO-US07115.
 XX
 PR 11-APR-1997; 97US-0041877.
 XX
 PA (WHEE) WHITEHEAD INST BIOMEDICAL RES.
 XX
 P1 Iahn BT, Page DC;
 XX
 DR WP1: 1998-568729/48.
 DR P-PSDB; W81501.
 XX
 PT Novel genes in the non-combining region of Y chromosome - useful to
 PT diagnose if male infertility or reduced sperm count has a genetic
 PT basis
 XX
 PS Disclosure: Fly 3A-B; 54pp; English.
 XX
 CC This represents the nucleotide sequence of the Dead Box X (DBX) gene long
 CC transcript. The invention relates to genes occurring in the non-
 CC recombining region of the human Y chromosome. The sequences fall into two
 CC classes: (1) X-homologous DNA which are expressed in many organs, having
 CC functional X homologies and (2) testis-specific DNA sequences. Y
 CC chromosomal DNA from males with known conditions such as infertility and
 CC reduced sperm count can be assessed using the invention to determine
 CC whether the condition is associated with or caused by the occurrence of
 CC the gene or gene alteration. Candidate inhibitors of the enzymatic
 CC activity of the genes can be assessed using in vitro assays.
 XX
 SQ Sequence 5322 BP; 1532 A; 1011 C; 1274 G; 1505 T; 0 other;

Query Match 89.3%; Score 13.4; DB 19; Length 5322;
 Best Local Similarity 93.3%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
 ||||| |||||
 Db 728 agcgggtctattaga 742

```

RESULT 6
081790/c
ID 081790 standard; DNA: 8957 BP.
AC 081790;
XX
DT 14-MAY-1996 (first entry)
XX
DE HTLV-II virus NRA strain genome sequence.
XX
KW HTLV-II; provirus; virus; NRA-19a; gag region; pol region;
KW env region; p19; p24; p15; gp46; p21e; tax/rex; tax region; p40x;
KW rex region; p26; cell culture; ATCC CRL 11580; fusion protein;
KW cloning; diagnostic; antibody; immunoassay; primer; probe;
KW ligase chain reaction; polymerase chain reaction; PCR;
KW Southern blot; hybridisation; Western blot; vaccine; ss.
XX
OS Human lymphotropic virus type II strain NRA-19a.
XX
PH Key
FT LTR
FT Location/Qualifiers
FT 1..766
FT /*tag- a
FT CDS 810..2111
FT /*tag- b
FT /*note= "gag region"
FT 810..1217
FT /*tag- c
FT /*product= p19 protein (R64971)
FT 1218..1859
FT /*tag- d
FT /*product= p24 protein (R64972)
FT 1860..2108
FT /*tag- e
FT /*product= p15 protein (R64973)
FT 2242..5190
FT /*tag= f
FT /*note= "pol region (including precursor)"
FT 2242..2436
FT /*tag- g
FT /*tag- h
FT /*product= Pol protein (R64974)
FT 5183..6643
FT /*tag- i
FT /*product= Env protein (R64975)
FT 6107..6640
FT /*note= "Env region"
FT mal_peptide
FT /*tag- j
FT /*product= p21e protein
FT 5183..8282
FT /*tag- k
FT /*product= p40x protein (R64976)
FT 5187..7215
FT /*note= "Tax region"
FT /*tag= l
FT CDS 5124..7665
FT /*tag- m
FT /*product= p26 protein (R64977)
FT /*note= "Rex region"
FT 5187..7215
FT /*tag- n
FT intron 8283..8957
FT LTR
FT /*tag- o
XX
MO9501457-A1.
XX
PD 12-JAN-1995.
XX
PF 30-JUN-1994; 94WO-US07419.
XX
PR 01-JUL-1993; 93US-0086415.
PR 20-JUN-1994; 94US-0259451.
XX

```

```

PA (ABBO ) ABBOTT LAB.
XX
PI Buytendorp MH, Chan EM, Chen ISY, Edwards M, Golde DM;
PI Guidinger P, Idler KB, Johnson JF, Lee HH, Motley CT;
PI Peterson B, Robertson E, Rosenblatt JD, Stephens JE;
PI Swanson PA, Tate C;
XX
DR WP1: 1995-061020/08.
DR P-PSDB: R64971-77.
XX
PT Human T-cell lymphotropic type II NRA viral genome, - expression
PT products, and compositions for use in improved methods for detection
PT of HTLV infection
XX
PS Claim 1: Page 39-45; 100pp; English.
XX
CC The complete sequence of the HTLV-II provirus NRA-19a strain genome is
CC presented. The gag region encodes proteins p19, p24 and p15, and
CC the pol region (including the pol precursor) and env region
CC (encoding gp46 and p21e) are indicated. The tax/rex region comprises
CC the tax region (encoding p40x) and the rex region (encoding p26),
CC which partially overlap, and are involved in regulating or controlling
CC gene expression. The virus may be produced in cell culture ATCC
CC CRL 11580, and the encoded proteins may be produced as recombinant
CC fusion proteins. The DNA and gene products may be used in
CC diagnostic methods for detection of HTLV-II DNA or antibodies in test
CC samples, by immunoassay, ligase chain reaction, polymerase chain
CC reaction, Southern blot hybridisation or Western blot methods, and in
CC antibody and vaccine preparation.
XX
SQ Sequence 8957 BP; 2154 A; 3208 C; 1648 G; 1947 T; 0 other;
XX
Query Match 89 38; Score 13.4; DB 16; Length 8957;
Best Local Similarity 93.38; Pred. No. 52;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 agcggatctattaga 15
Db 6653 AGCGTGTCTATTATA 6639
XX
RESULT 7
ID Q05577 standard; DNA: 360 BP.
XX
AC Q05577;
XX
DT 19-JHC 1990 (first entry)
XX
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
DE antigen polypeptide.
XX
KW Mycoplasma pneumoniae; enzootic pneumonia; ds;
XX
OS Mycoplasma hyopneumoniae.
XX
PN JP02167079-A.
XX
PD 27-JUN-1990.
XX
PF 21-DEC-1988; 88JP-0322829.
XX
PR 21-DEC-1988; 88JP-0322829.
XX
PA (NISE-) NIPPON SEIPUN KK.
XX
DR WP1: 1990-241949/32.
DR P-PSDB: R06278.
XX
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs

```

XX Claim 1: Page 570: 28pp; Japanese.
 XX Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 XX
 XX Sequence 360 BP; 130 A; 52 C; 64 G; 114 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 360;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
 |||
 Db 61 gcgggtctattag 73

RESULT 8
 Q05576
 ID Q05576 standard; DNA; 396 BP.
 AC Q05576;
 XX
 XX 19-DEC-1990 (first entry)
 DT
 XX Sequence encoding swine enzootic pneumonia mycoplasma surface
 DE antigen polypeptide.
 DE
 XX Mycoplasma pneumonia; enzootic pneumonia; ds;
 KW Mycoplasma hyopneumoniae.
 OS
 XX JP02167079-A.
 PN
 XX 27-JUN-1990.
 PD
 XX 21-DEC-1988; 88JP-0322829.
 PF
 XX 21-DEC-1988; 88JP-0322829.
 PR
 XX (NLSB-) NIPPON SEIFUN KK.
 PA
 DR WPI: 1990-241949/32.
 DR P-PSDB; R06277.
 DR
 XX Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
 PT for early detection and treatment of mycoplasma or enzootic
 PT pneumonia of pigs
 PT
 XX Claim 2: Page 570: 28pp; Japanese.
 PS
 XX Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 CC
 XX Sequence 396 BP; 145 A; 62 C; 70 G; 119 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 396;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
 |||
 Db 43 gcgggtctattag 55

RESULT 9
 Q05578
 ID Q05578 standard; DNA; 843 BP.
 XX

AC Q05578;
 XX
 XX 19-DEC-1990 (first entry)
 DT
 XX Sequence encoding swine enzootic pneumonia mycoplasma surface
 DE antigen polypeptide.
 DE
 XX Mycoplasma pneumonia; enzootic pneumonia; ds;
 KW Mycoplasma hyopneumoniae.
 OS
 XX JP02167079-A.
 PN
 XX 27-JUN-1990.
 PD
 XX 21-DEC-1988; 88JP-0322829.
 PF
 XX 21-DEC-1988; 88JP-0322829.
 PR
 XX 21-DEC-1988; 88JP-0322829.
 XX
 XX (NLSB-) NIPPON SEIFUN KK.
 PA
 DR WPI: 1990-241949/32.
 DR P-PSDB; R06279.
 DR
 XX Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
 PT for early detection and treatment of mycoplasma or enzootic
 PT pneumonia of pigs
 PT
 XX Claim 2: Page 570: 28pp; Japanese.
 PS
 XX Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 CC
 XX Sequence 843 BP; 324 A; 127 C; 139 G; 253 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 843;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
 |||
 Db 103 gcgggtctattag 115

RESULT 10
 F13364/c
 ID F13364 standard; cDNA; 1151 BP.
 XX
 XX F13364;
 AC
 XX 13-MAR-2001 (first entry)
 DT
 XX Aspergillus oryzae RST SEQ ID NO:5887.
 DE
 XX
 XX Multiple gene expression; filamentous fungal cell; RST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Aspergillus oryzae.
 OS
 XX WO200056762-A2.
 PN
 XX 28-SEP-2000.
 PD
 XX 22-MAR-2000; 2000WO-US07781.
 PF
 XX 22-MAR-1999; 99US-0273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA

XX Doughty SW, Lee R, Walker J;
 XX WPI: 1996-433763/43.
 DR P-PSDB; W01037.
 XX
 PT Putative protective antigens against Mycoplasma - used for the
 PT detection, prevention or treatment of Mycoplasma infections, esp. M.
 PT hyopneumoniae in swine
 XX
 PS Claim 14: Page 28-29; 43pp; English.
 XX
 CC The gene (T38241) coding for a 48 kDa putative protective antigen
 CC (W01037) was obt'd. from a Mycoplasma hyopneumoniae genomic library
 CC by screening with a probe generated by PCR amplification (see also
 CC T13814-16). The antigen had originally been isolated from M.
 CC hyopneumoniae cells using antibody probes enriched with Mycoplasma-
 CC specific antibodies. Other protective antigens were also identified
 CC (see also W01024-31). Protective antigens and antibodies can be
 CC used in vaccines for preventing or treating mycoplasma infections,
 CC partic. M. hyopneumoniae infections in swine. They can also be used
 CC for diagnosis.
 CC
 SQ Sequence 1782 BP; 682 A; 254 C; 259 G; 587 T; 0 other;
 XX
 OY 2 gcgggtctattag 14
 Db 1039 gcgggtctattag 1051
 XX
 RESULT 13
 X21072
 ID X21072 standard; DNA; 533 BP.
 XX
 AC X21072;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN W09859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PE 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI: 1999-081273/07.
 XX
 PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1; Page 1052; 1150pp; English.
 XX
 CC X20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum

CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 533 BP; 131 A; 111 C; 145 G; 143 T; 3 other;
 XX
 OY 2 gcgggtctattag 15
 Db 1 gcgggtctattag 14
 XX
 RESULT 14
 T79770
 ID T79770 standard; DNA; 1167 BP.
 XX
 AC T79770;
 XX
 DT 10-NOV-1997 (first entry)
 XX
 DE Potato starch branching enzyme gene intron 1.
 XX
 KW Starch branching enzyme; SBE; potato; genetic engineering;
 KW transgenic plant; starch production; modification; ss.
 XX
 OS Solanum tuberosum.
 XX
 PN W09704113-A2.
 XX
 PD 06-FEB-1997.
 XX
 PE 12-JUL-1996; 96WO-EP03053.
 XX
 PR 14-JUL-1995; 95GB-0014437.
 XX
 PA (DANT-) DANISCO AS.
 XX
 PI Poulsen P;
 XX
 DR WPI: 1997-132651/12.
 XX
 PT Affecting enzymatic activity in plant by expressing sense intron
 PT sequence - esp. for inhibiting starch branching enzyme to create
 PT modified starch without post-harvest derivatisation
 XX
 PS Claim 8; Page 30; 70pp; English.
 XX
 CC T79770 represents intron 1 of the potato starch-branching enzyme
 CC (SBE) gene. Sequences antisense to introns of the SBE gene are used
 CC to produce transgenic plants producing modified or new forms of starch
 CC for industrial use. The sequences allow starches to be made without
 CC post-harvest modification (which uses hazardous chemicals and requires
 CC much energy). Using the antisense intron sequences is more reliable,
 CC efficient and selective than known methods for regulating enzymatic
 CC activity because there is almost no homology between introns in
 CC different SBE genes. The SBE gene promoter (see T79783) is also
 CC useful in genetic engineering.
 XX
 SQ Sequence 1167 BP; 324 A; 155 C; 228 G; 458 T; 2 other;
 XX
 OY 2 gcgggtctattag 15
 Db 537 gcgggtctattag 550
 XX
 Query Match 82.7%; Score 12.4; DB 20; Length 533;
 Best Local Similarity 92.9%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

RESULT 15
T79791/c
ID T79791 standard; DNA; 1167 bp.
XX
AC T79791.
XX
DT 11-NOV-1997 (first entry)
XX
DE potato starch branching enzyme gene intron 1 antisense sequence.
XX
KM Starch branching enzyme; SBE; potato; genetic engineering;
XX transgenic plant; starch production; modification; ss.
XX
OS Solanum tuberosum.
XX
PN M09704112-A2.
XX
PD 06-FEB-1997.
XX
PF 12-JUL-1996; 96WO-EP03052.
XX
PR 14-JUL-1995; 95GB-0014435.
XX
PA (DAMI-) DANISCO AS.
XX
PI Poulsen P.
XX
DR WP1; 1997-132650/12.
XX
PT Affecting enzymatic activity in plant by expressing anti-sense
PT intron sequence esp. for inhibiting starch branching enzyme thus
PT creating modified starch without post-harvest derivatisation
XX
PS Claim 8; Page 38; 70pp; English.
XX
CC T79791 is an antisense sequence of intron 1 of the potato starch-
CC branching enzyme (SBE) gene. Sequences antisense to introns of the
CC SBE gene were used to produce transgenic plants producing modified
CC or new forms of starch for industrial use. The sequences allow
CC starches to be made without post-harvest modification (which uses
CC hazardous chemicals and requires much energy). Using the antisense
CC intron sequences is more reliable, efficient and selective than
CC known methods for regulating enzymatic activity because there is
CC almost no homology between introns in different SBE genes. The SBE
CC gene promoter (see T79804) is also useful in genetic engineering.
XX
SQ Sequence 1167 bp; 458 A; 228 C; 155 G; 324 T; 2 other;

```

```

Query Match      82.7%; Score 12.4; HH 18; Length 1167;
Best Local Similarity 92.9%; Pred. No. 1.8e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 qcgggtctattaga 15
   | ||||| |||||
Db 631 GAGGCTGTATTAGA 618

```

Search completed: May 1, 2001, 18:20:18
Job time: 5909 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 15:51:57 ; Search time 65.16 Seconds
(without alignments)

40.194 Million cell updates/sec

Title: US-09-593-914-1

Perfect score: 15

Sequence: 1 agcgggtctattaga 15

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : Issued Parents NA:*

1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*

2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*

3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*

4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*

5: /cgn2_6/prodata/1/ina/PCITUS.COMB.seq:*

6: /cgn2_6/prodata/1/ina/backfill1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13.4	89.3	3408	4	US-09-058-489-14
2	13.4	89.3	5322	4	US-09-058-489-13
3	12.4	82.7	1167	4	US-08-981-803-1
4	12.4	82.7	1167	4	US-08-981-803-15
5	12.4	82.7	2022	2	US-08-505-486-96
6	12.4	82.7	2022	3	US-08-801-028-96
7	12.4	82.7	2022	3	US-09-340-154-95
8	12.4	82.7	2022	5	US-09-340-154-96
9	12.4	82.7	2022	5	PCM-US95-09338-96
10	12.4	82.7	2127	2	US-08-505-486-95
11	12.4	82.7	2127	3	US-08-801-028-95
12	12.4	82.7	2127	3	US-09-340-154-95
13	12.4	82.7	2127	5	PCM-US95-09338-95
14	12.4	82.7	2250	5	PCM-US95-09339-95
15	12.4	82.7	2250	5	PCT-US95-09339-95
16	12.4	82.7	11478	4	US-08-981-803-29
17	11.8	78.7	68	2	US-08-984-578-151
18	11.8	78.7	347	1	US-08-149-695-3
19	11.8	78.7	347	1	US-08-377-228-3
20	11.8	78.7	377	1	US-08-149-695-4
21	11.8	78.7	377	1	US-08-377-228-4
22	11.8	78.7	464	1	US-08-149-695-5
23	11.8	78.7	464	1	US-08-377-228-5
24	11.8	78.7	595	1	US-08-149-695-6
25	11.8	78.7	595	1	US-08-377-228-6
26	11.8	78.7	735	3	US-09-135-782-3
27	11.8	78.7	1000	1	US-08-117-083-63

28	11.8	78.7	1579	1	US-08-149-695-7	Sequence 7, Appl1
29	11.8	78.7	1579	1	US-08-377-228-7	Sequence 7, Appl1
30	11.8	78.7	1835	1	US-08-361-611-3	Sequence 3, Appl1
31	11.8	78.7	1835	1	US-08-365-655-3	Sequence 3, Appl1
32	11.8	78.7	1835	2	US-08-946-967-3	Sequence 3, Appl1
33	11.8	78.7	2454	3	US-09-221-235-7	Sequence 7, Appl1
34	11.8	78.7	2454	4	US-09-221-235-7	Sequence 7, Appl1
35	11.8	78.7	2454	4	US-09-221-527-7	Sequence 7, Appl1
36	11.8	78.7	2454	4	US-09-221-236-7	Sequence 7, Appl1
37	11.8	78.7	2454	4	US-09-221-416-7	Sequence 7, Appl1
38	11.8	78.7	2454	4	US-09-221-245-7	Sequence 7, Appl1
39	11.8	78.7	2454	4	US-09-163-115-7	Sequence 7, Appl1
40	11.8	78.7	2454	4	US-09-221-528-7	Sequence 7, Appl1
41	11.8	78.7	2454	4	US-09-593-553-7	Sequence 7, Appl1
42	11.8	78.7	2687	1	US-08-149-695-8	Sequence 8, Appl1
43	11.8	78.7	2687	1	US-08-377-228-8	Sequence 8, Appl1
44	11.8	78.7	2873	1	US-08-149-695-1	Sequence 1, Appl1
45	11.8	78.7	2873	1	US-08-377-228-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1

US-09-058-489-14

Sequence 14, Application US/09058489

Patent No. 6103886

GENERAL INFORMATION:

APPLICANT: Whitehead Institute for Biomedical Research

APPLICANT: Lahn, Bruce

APPLICANT: Page, David

TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

FILE REFERENCE: WH197-08PA

CURRENT APPLICATION NUMBER: US/09/058,489

CURRENT FILING DATE: 1998-04-10

EARLIER APPLICATION NUMBER: 60/041,877

EARLIER FILING DATE: 1997-04-11

NUMBER OF SEQ ID NOS: 91

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 14

LENGTH: 3408

TYPE: DNA

ORGANISM: Human

US-09-058-489-14

Query Match

Best local similarity 93.3%; Score 13.4; DB 4; Length 3408;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 728 agcgggtctattaga 742

RESULT 2

US-09-058-489-13

Sequence 13, Application US/09058489

Patent No. 6103886

GENERAL INFORMATION:

APPLICANT: Whitehead Institute for Biomedical Research

APPLICANT: Lahn, Bruce

APPLICANT: Page, David

TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

FILE REFERENCE: WH197-08PA

CURRENT APPLICATION NUMBER: US/09/058,489

CURRENT FILING DATE: 1998-04-10

EARLIER APPLICATION NUMBER: 60/041,877

EARLIER FILING DATE: 1997-04-11

NUMBER OF SEQ ID NOS: 91

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 13
LENGTH: 5322
TYPE: DNA
ORGANISM: Human
US-09-058-489-13

Query Match 89.3%; Score 13.4; DB 4; Length 5322;
Best Local Similarity 93.3%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 accgagctattaga 15
|||||
Db 728 accgagctattaga 742

RESULT 3
US-08-981-803-1
Sequence 1, Application US/08981803
Patent No. 6147279
GENERAL INFORMATION:
APPLICANT: POULSEN, PETER
TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION
FILE REFERENCE: 674509-2004
CURRENT APPLICATION NUMBER: US/08/981,803
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: PCT/EP96/03052
EARLIER FILING DATE: 1996-07-12
EARLIER APPLICATION NUMBER: 9514435.8
EARLIER FILING DATE: 1995-07-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1167
TYPE: DNA
ORGANISM: Solanum tuberosum
US-08-981-803-1

Query Match 82.7%; Score 12.4; DB 4; Length 1167;
Best Local Similarity 92.9%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaggctattaga 15
|||||
Db 537 gaggctattaga 550

RESULT 4
US-08-981-803-15/C
Sequence 15, Application US/08981803
Patent No. 6147279
GENERAL INFORMATION:
APPLICANT: POULSEN, PETER
TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION
FILE REFERENCE: 674509-2004
CURRENT APPLICATION NUMBER: US/08/981,803
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: PCT/EP96/03052
EARLIER FILING DATE: 1996-07-12
EARLIER APPLICATION NUMBER: 9514435.8
EARLIER FILING DATE: 1995-07-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 1167
TYPE: DNA
ORGANISM: Solanum tuberosum
US-08-981-803-15

Query Match 82.7%; Score 12.4; DB 4; Length 1167;
Best Local Similarity 92.9%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gaggctattaga 15
|||||
Db 631 GAGGCTATTAGA 618

RESULT 5
US-08-505-486-96/C
Sequence 96, Application US/08505486
Patent No. 5955573
GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROTIMBEL, FIGG, ERNST & KURZ
STREET: 555 Thirteenth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,486
FILING DATE: 21-JUL-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 2022
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULAR TYPE:
DESCRIPTION: CHROMIC DNA
US-08-505-486-96

Query Match 82.7%; Score 12.4; DB 2; Length 2022;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaggctattaga 15
|||||
Db 647 GAGGCTATTAGA 634

RESULT 6
US-08-801-028-96/C
Sequence 96, Application US/08801028
Patent No. 6018102
GENERAL INFORMATION:
APPLICANT: JOAN GARBARTINO
APPLICANT: JESSE M. JAYNES
APPLICANT: WILLIAM BELKNAP
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN

```

1  COUNTRY: USA
2  ZIP: 20004
3  COMPUTER READABLE FORM:
4  MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
5  COMPUTER: IBM COMPATIBLE
6  OPERATING SYSTEM: DOS
7  SOFTWARE: Wordperfect 5.1+
8  CURRENT APPLICATION DATA:
9  APPLICATION NUMBER: US/09/340,154
10 FILING DATE:
11 CLASSIFICATION:
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: 08/505,486
14 FILING DATE: 21-JUL-1995
15 APPLICATION NUMBER: U.S. 08/279,472
16 FILING DATE: 22-JUL-1994
17 ATTORNEY/AGENT INFORMATION:
18 NAME: WALKER, BARBARA W.
19 REGISTRATION NUMBER: 35,400
20 REFERENCE/DOCK#T NUMBER: 2093-117A
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (202)783-6040
23 TELEFAX: (202)783-6031
24 INFORMATION FOR SEQ ID NO: 96:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 2022
27 TYPE: NUCLEIC ACID
28 STRANDEDNESS: DOUBLE STRANDED
29 TOPOLOGY: LINEAR
30 MOLECULE TYPE:
31 DESCRIPTION: GENOMIC DNA
32 US-09-340-154-96
33
34 Query Match 82.7%; Score 12.4; DB 3; Length 2022;
35 Best local similarity 92.9%; Pred. No. 55;
36 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
37
38 QY 2 gggggctcattaga 15
39 1 | | | | | | | | | |
40 DB 647 GAGGGTCTATTAGA 634
41
42 RESULT 8
43 PCT-US95-09338-96/C
44 Sequence 96, Application PC/TUS9509338
45 GENERAL INFORMATION:
46 APPLICANT:
47 TITLE OF INVENTION: UR10UPT1N-1YTIC PEPTIDE FUSION GENE
48 TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
49 NUMBER OF SEQUENCES: 98
50 COMPUTER READABLE FORM:
51 MEDIUM TYPE: Floppy disk
52 COMPUTER: IBM PC compatible
53 OPERATING SYSTEM: PC-DOS/MS-DOS
54 SOFTWARE: WORDPERFECT 5.1+
55 CURRENT APPLICATION DATA:
56 APPLICATION NUMBER: PCT/US95/09338
57 FILING DATE: 21-JUL-1994
58 PRIOR APPLICATION DATA: 08/279,472
59 APPLICATION NUMBER: 08/279,472
60 FILING DATE: 22-JUL-1994
61 INFORMATION FOR SEQ ID NO: 96:
62 SEQUENCE CHARACTERISTICS:
63 LENGTH: 2022
64 TYPE: NUCLEIC ACID
65 STRANDEDNESS: DOUBLE STRANDED
66 TOPOLOGY: LINEAR
67 MOLECULE TYPE:
68 DESCRIPTION: GENOMIC DNA
69 PCT-US95-09338-96

```


APPLICATION NUMBER: 08/148,889
FILING DATE: 11-08-93
ATTORNEY/AGENT INFORMATION:
NAME: WASSERMAN, FRANK S.
REGISTRATION NUMBER: 34,273
REFERENCE/DOCKET NUMBER: 4013-104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)990-9531
TELEFAX: (919)990-9532
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-08-801-028-95

Query Match 82.7%; Score 12.4; DB 3; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcgggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 12
US-09-340-154-95/c
Sequence 95, Application US/09340154
Patent No. 6084156
GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESS: ROTHWELL, FIGG, ERNST & KURZ
STREET: 555 Thirteenth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1+
CURRENT APPLICATION DATA: 09/340.154
APPLICATION NUMBER: US/09/340.154
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/505,486
FILING DATE: 21-JUL-1995
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR

MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-09-340-154-95

Query Match 82.7%; Score 12.4; DB 3; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcgggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 13
PCT-US95-09338-95/c
Sequence 95, Application PCT/US9509338
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
NUMBER OF SEQUENCES: 98
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09338
FILING DATE: 21-JUL-1994
PRIOR APPLICATION DATA: 08/279,472
APPLICATION NUMBER: 08/279,472
FILING DATE: 22-JUL-1994
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
PCT-US95-09338-95

Query Match 82.7%; Score 12.4; DB 5; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcgggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 14
PCT-US95-09339-95/c
Sequence 95, Application PCT/US9509339
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
NUMBER OF SEQUENCES: 98
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09339
FILING DATE: 21-JUL-1994
PRIOR APPLICATION DATA:

Thu May 3 07:47:08 2001

us-09-593-914-1.rni

Page 7

117: qb_est148:*
 118: qb_est149:*
 119: qb_est150:*
 120: qb_est151:*
 121: qb_est152:*
 122: qb_est153:*
 123: qb_est154:*
 124: qb_est155:*
 125: qb_est156:*
 126: qb_est157:*
 127: qb_est158:*
 128: qb_est159:*
 129: qb_est160:*
 130: qb_est161:*
 131: qb_est162:*
 132: qb_est163:*
 133: qb_est164:*
 134: qb_est165:*
 135: qb_est166:*
 136: qb_est167:*
 137: qb_est168:*
 138: qb_est169:*
 139: qb_est170:*
 140: qb_est171:*
 141: qb_est172:*
 142: qb_est173:*
 143: qb_est174:*
 144: qb_est175:*
 145: qb_est176:*
 146: qb_est177:*
 147: qb_est178:*
 148: qb_est179:*
 149: qb_est180:*
 150: qb_est181:*
 151: qb_est182:*
 152: qb_est183:*
 153: qb_est184:*
 154: qb_est185:*
 155: qb_est186:*
 156: qb_est187:*
 157: qb_est188:*
 158: qb_est189:*
 159: qb_est190:*
 160: qb_est191:*
 161: qb_est192:*
 162: qb_est193:*
 163: qb_est194:*
 164: qb_est195:*
 165: qb_est196:*
 166: qb_est197:*
 167: qb_est198:*
 168: qb_est199:*
 169: qb_est200:*
 170: qb_est201:*
 171: qb_est202:*
 172: qb_est203:*
 173: qb_est204:*
 174: qb_est205:*
 175: qb_est206:*
 176: qb_est207:*
 177: qb_est208:*
 178: qb_est209:*
 179: qb_est210:*
 180: qb_est211:*
 181: qb_est212:*
 182: qb_est213:*
 183: qb_est214:*
 184: qb_est215:*
 185: qb_est216:*
 186: qb_est217:*
 187: qb_est218:*
 188: qb_est219:*
 189: qb_est220:*

190: em_gss_pln1:*
 191: em_gss_pln2:*
 192: em_gss_pro:*
 193: em_gss_rtd1:*
 194: em_gss_rtd2:*
 195: em_gss_rtd3:*
 196: em_gss_rtd4:*
 197: em_gss_rtd5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: qb_gss1:*
 202: qb_gss2:*
 203: qb_gss3:*
 204: qb_gss4:*
 205: qb_gss5:*
 206: qb_gss6:*
 207: qb_gss7:*
 208: qb_gss8:*
 209: qb_gss9:*
 210: qb_gss10:*
 211: qb_gss11:*
 212: qb_gss12:*
 213: qb_gss13:*
 214: qb_gss14:*
 215: qb_gss15:*
 216: qb_gss16:*
 217: qb_gss17:*
 218: qb_gss18:*
 219: qb_gss19:*
 220: qb_gss20:*
 221: qb_gss21:*
 222: qb_gss22:*
 223: qb_gss23:*
 224: qb_gss24:*
 225: qb_gss25:*
 226: qb_gss26:*
 227: qb_gss27:*
 228: qb_gss28:*
 229: qb_gss29:*
 230: qb_gss30:*
 231: qb_gss31:*
 232: qb_gss32:*
 233: qb_gss33:*
 234: qb_gss34:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	446	211	A0774045	A0774045 HS_2190_B
2	14	93.3	171	226	H000027	B000027 CSRL-100E3-
3	14	93.3	412	203	A0166430	A0166430 HS_3128_B
4	14	93.3	570	167	B0436368	B0436368 EST407446
5	14	93.3	616	144	B0113023	B0113023 EST440533
6	13.4	89.3	146	119	AW723826	AW723826 f1050m.f
7	13.4	89.3	149	18	A1320373	A1320373 c3050m.f
8	13.4	89.3	151	119	AW724282	AW724282 f4906m.f
9	13.4	89.3	151	119	AW724283	AW724283 f4906m.f
10	13.4	89.3	187	119	AW715484	AW715484 g5c070m.f
11	13.4	89.3	189	18	A1321118	A1321118 g4d10cm.f
12	13.4	89.3	201	133	B0429584	B0429584 B0429584
13	13.4	89.3	308	127	B0200417	B0200417 B0200417
14	13.4	89.3	316	201	A0072101	A0072101 HS_3010_B
15	13.4	89.3	323	18	A1318822	A1318822 a21030m.f
16	13.4	89.3	326	111	AW171957	AW171957 618056E01
17	13.4	89.3	368	218	A2267638	A2267638 RPCI-23-1
18	13.4	89.3	391	1	AA003631	AA003631 mg60c05.r

```

C 19 13.4 89.3 392 18 A1320433 A1320433 c9g01nm.f
C 20 13.4 89.3 392 18 A1321592 A1321592 d9g12nm.f
C 21 13.4 89.3 392 18 A1321668 A1321668 e1q05nm.f
C 22 13.4 89.3 392 119 A13218206 A13218206 f14cl0nm.f
C 23 13.4 89.3 392 18 A1321689 A1321689 e2a10nm.f
C 24 13.4 89.3 392 119 A13215283 A13215283 g2e01nm.f
C 25 13.4 89.3 392 119 A13218413 A13218413 f5c02nm.f
C 26 13.4 89.3 392 119 A13218484 A13218484 f8e09nm.f
C 27 13.4 89.3 392 119 A1321961 A1321961 f9b09nm.f
C 28 13.4 89.3 392 119 A1321962 A1321962 f9b09nm.f
C 29 13.4 89.3 425 119 A13218414 A13218414 f5c02nm.f
C 30 13.4 89.3 425 23 A13218709 A13218709 f5c02nm.f
C 31 13.4 89.3 425 18 A13218717 A13218717 f5c02nm.f
C 32 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 33 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 34 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 35 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 36 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 37 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 38 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 39 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 40 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 41 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 42 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 43 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 44 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f
C 45 13.4 89.3 425 119 A13218717 A13218717 f5c02nm.f

```

ALIGNMENTS

```

RESULT 1
LOCUS A0774045 446 bp DNA
DEFINITION HS.2190.B2.H07.WTC.C17 Approved Human Genomic Sperm Library D Homo
ACCESSION A0774045
VERSION A0774045.1 GI:5653773
KEYWORDS GSS.
ORGANISM Homo sapiens
SOURCE human.

```

```

REFERENCE 1
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2190 row: P column: 14
Seq primer: 17
Class: BAC ends
High quality sequence stop: 446.
Location/Qualifiers
1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-2190 COL-14 Row-P"
/clone_lib="C17 Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: peltOBAC11; BAC clones in

```

FEATURES

```

source
1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-2190 COL-14 Row-P"
/clone_lib="C17 Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: peltOBAC11; BAC clones in

```

```

BASE COUNT 160 a 72 c 102 g 110 t 2 others
ORIGIN

```

```

Query Match 100.0%; Score 15; DB 211; Length 446;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 aacgggctcattaga 15
Db 202 ACCGGCTCTATTAGA 155

```

```

RESULT 2
LOCUS B00027 171 bp DNA
DEFINITION CSRL-100c3-u CSRL flow sorted Chromosome 11 specific cosmid Homo
ACCESSION B00027
VERSION B00027.1 GI:1409305
KEYWORDS GSS.
ORGANISM Homo sapiens
SOURCE human.

```

```

REFERENCE 1
AUTHORS Evans,G.A., Burdee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S., Harris,
J., Bedford,J., McFarland,J., Butzinski,K., Kluen,M., Kupfer,K. and
Garner,H.K.
Genomic Sequence Sampled Map of Chromosome 11
Unpublished (1996)
Contact: Evans GA, Shane Probst
Molecular Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: geyans@utsw.swmed.edu, shane@mdermoll.swmed.edu
Seq primer: 17
Class: cosmid ends
High quality sequence stop: 171.
Location/Qualifiers
1..171
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-100c3"
/clone_lib="CSRL flow sorted Chromosome 11 specific
cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid
library prepared from flow sorted human Chromosome 11
derived from Chinese Hamster Ovary (CHO) monochromosomal
somatic cell hybrid, J1"

```

FEATURES

```

source
1..171
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-100c3"
/clone_lib="CSRL flow sorted Chromosome 11 specific
cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid
library prepared from flow sorted human Chromosome 11
derived from Chinese Hamster Ovary (CHO) monochromosomal
somatic cell hybrid, J1"
BASE COUNT 49 a 42 c 38 g 34 t 8 others
ORIGIN

```

```

Query Match 93.3%; Score 14; DB 226; Length 171;
Best Local Similarity 93.3%; Pred. No. 1,8e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 aacgggctcattaga 15
Db 169 AAGGGCTCTATTAGA 155

```

```

RESULT 3
LOCUS A0166430 412 bp DNA
DEFINITION HS_3128_B1_G03_MR C17 Approved Human Genomic Sperm Library D Homo

```

accession sapiens genome clone Plate-3128 Col-5 Row-N, DNA sequence.
 version AQ166430
 keywords AQ166430.1 GI:3564625
 source GSS.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 412)
 reference Maitras,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 authors Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
 Hood,L.
 title Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 journal Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 medline 99380589
 comment Contact: Maitras G., Wallace J.C., Hood L.
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 sequence tagged connector
 plate: 3128 row: N column: 5
 class: BAC ends
 high quality sequence stop: 412.
 location/qualifiers
 1..412
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="3128 Col-5 Row-N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pELODACL1; BAC clones in
 E-Coli DH10B"
 base count 108 a 84 c 77 g 143 t
 origin
 Query Match 93.3%; Score 14; DB 203; Length 412;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agggggtctatag 14
 |||
 Db 356 AGCGGCTATTAG 369
 RESULT 4
 BE436368 570 bp mRNA KSP 24-JUL-2000
 DEFINITION EST407446 tomato breaker fruit, "TIGR Lycopersicon esculentum cDNA
 accession BE436368
 version BE436368.1 GI:9434211
 keywords EST.
 source tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids
 1 (bases 1 to 570)
 reference Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
 authors Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
 S.D.
 title Generation of ESTs from tomato fruit tissue, breaker stage
 journal Unpublished (2000)
 comment Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfisch@clemson.edu
 5 prime sequence.
 location/qualifiers
 1..570
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone_plate="3128 Col-5 Row-N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /issue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmCudapt; Site.1: EcoRI;
 site.2: XhoI; Fruit were harvested at the breaker stage
 (first sign of lycopene accumulation on the blossom end of
 the fruit). Fruit were cut in half and the seeds and
 locules were discarded prior to freezing the pericarp."
 base count 186 a 128 c 83 g 173 t
 origin
 Query Match 93.3%; Score 14; DB 167; Length 570;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agggggtctatag 14
 |||
 Db 241 AGCGGCTATTAG 254
 RESULT 5
 BE113023 616 bp mRNA EST 20-OCT-2000
 DEFINITION EST440533 tomato breaker fruit Lycopersicon esculentum cDNA clone
 accession BE113023
 version BE113023
 keywords EST.
 source tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids
 1 (bases 1 to 616)
 reference Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
 authors Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
 S.D.
 title Generation of ESTs from tomato fruit tissue, breaker stage
 journal Unpublished (2000)
 comment Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfisch@clemson.edu
 location/qualifiers
 1..616
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone_plate="3128 Col-5 Row-N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /issue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmCudapt; Site.1: EcoRI;
 site.2: XhoI; Fruit were harvested at the breaker stage
 (first sign of lycopene accumulation on the blossom end of
 the fruit). Fruit were cut in half and the seeds and

locules were discarded prior to freezing the pericarp."

BASE COUNT 199 a 138 c 90 g 189 t

ORIGIN

Query Match 93.3%; Score 14; DB 144; Length 616;
 Best Local Similarity 100.0%; Pred. No. 1,9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 agcggctcattag 14
 |||
 Db 245 ACCGGGCTCTATTAG 258

RESULT 6
 AW723826 146 bp mRNA EST 19-APR-2000
 LOCUS f1f05nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
 DEFINITION cDNA clone f1f05nm 5', mRNA sequence.
 ACCESSION AW723826
 VERSION AW723826.1 GI:7618386
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 146)
 Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
 Two Neurospora crassa EST Databases
 Unpublished (1998)

COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: Universal Forward Primer
 High quality sequence slot: 121.

FEATURES
 Location/Qualifiers

1..146
 /organism="Neurospora crassa"
 /strain="bd, frq7 A"
 /db_xref="taxon:5141"
 /clone="f1f05nm"
 /clone_lib="Neurospora crassa morning cDNA library"
 /tissue_type="tissue harvested following 22hr growth in dark"
 /note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."
 BASE COUNT 58 a 26 c 18 g 44 t
 ORIGIN

Query Match 89.3%; Score 13.4; DB 119; Length 146;
 Best Local Similarity 93.3%; Pred. No. 4.2e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 agcggctcattag 15
 |||
 Db 49 AGCGGCTCTATTATA 35

RESULT 7
 A1320373 149 bp mRNA EST 18-DEC-1998
 LOCUS c3h05nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
 DEFINITION cDNA clone c3h05nm 5', mRNA sequence.
 ACCESSION A1320373

VERSION A1320373.1 GI:4036355
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa.

REFERENCE 1 (bases 1 to 149)
 Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
 Two Neurospora crassa EST Databases
 Unpublished (1998)
 Other ESTs: c3h05nm.f1
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: Universal Forward Primer
 High quality sequence slot: 127.

FEATURES
 Location/Qualifiers

1..149
 /organism="Neurospora crassa"
 /strain="bd, frq7 A"
 /db_xref="taxon:5141"
 /clone="c3h05nm"
 /clone_lib="Neurospora crassa morning cDNA library"
 /tissue_type="tissue harvested following 22hr growth in dark"
 /note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."
 BASE COUNT 59 a 26 c 19 g 45 t
 ORIGIN

Query Match 89.3%; Score 13.4; DB 18; Length 149;
 Best Local Similarity 93.3%; Pred. No. 4.3e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 agcggctcattag 15
 |||
 Db 49 AGCGGCTCTATTATA 35

RESULT 8
 AW724282 151 bp mRNA EST 19-APR-2000
 LOCUS f1f06nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
 DEFINITION cDNA clone f1f06nm 5', mRNA sequence.
 ACCESSION AW724282
 VERSION AW724282.1 GI:7618842
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa.

REFERENCE 1 (bases 1 to 151)
 Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
 Two Neurospora crassa EST Databases
 Unpublished (1998)
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center

Seq primer: Universal Forward Primer
High quality sequence stop: 116.
Location/Qualifiers
1. 151

FEATURES

source

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="f4g06nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

BASE COUNT
59 a 26 c 21 g 45 t
ORIGIN

Query Match
Best Local Similarity 93.3%; Score 13.4; DB 119; Length 151;
Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
|||||
Db 49 AGCGGCTATTATA 35

AM724283 151 bp mRNA EST 19-APR-2000
LOCUS f4g06nm.f2 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION CNA clone f4g06nm 5', mRNA sequence.
AM724283
VERSION AM724283.1 GI:7618843
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 151)
Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
Two Neurospora crassa EST databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 116.
Location/Qualifiers
1. 151

FEATURES

source

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="f4g06nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

BASE COUNT
59 a 26 c 21 g 45 t
ORIGIN

Query Match
Best Local Similarity 93.3%; Score 13.4; DB 119; Length 151;
Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
|||||
Db 49 AGCGGCTATTATA 35

AM724283 151 bp mRNA EST 19-APR-2000
LOCUS f4g06nm.f2 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION CNA clone f4g06nm 5', mRNA sequence.
AM724283
VERSION AM724283.1 GI:7618843
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 151)
Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
Two Neurospora crassa EST databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 116.
Location/Qualifiers
1. 151

BASE COUNT
59 a 26 c 21 g 45 t
ORIGIN

Query Match
89.3%; Score 13.4; DB 119; Length 151;

Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
|||||
Db 49 AGCGGCTATTATA 35

AM715484 187 bp mRNA EST 19-APR-2000
LOCUS 95c07nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION CNA clone 95c07nm 5', mRNA sequence.
AM715484
VERSION AM715484.1 GI:7604747
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 187)
Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
Two Neurospora crassa EST databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 152.
Location/Qualifiers
1. 187

FEATURES

source

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="95c07nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

BASE COUNT
69 a 35 c 23 g 60 t
ORIGIN

Query Match
Best Local Similarity 93.3%; Score 13.4; DB 119; Length 187;
Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
|||||
Db 49 AGCGGCTATTATA 35

AM715484 187 bp mRNA EST 18-DEC-1998
LOCUS d4d10nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION CNA clone d4d10nm 5', mRNA sequence.
AM715484
VERSION AM715484.1 GI:4037100
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 187)
Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
Two Neurospora crassa EST databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 152.
Location/Qualifiers
1. 187

BASE COUNT
69 a 35 c 23 g 60 t
ORIGIN

Query Match
Best Local Similarity 93.3%; Score 13.4; DB 119; Length 187;
Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
|||||
Db 49 AGCGGCTATTATA 35

AM715484 187 bp mRNA EST 18-DEC-1998
LOCUS d4d10nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION CNA clone d4d10nm 5', mRNA sequence.
AM715484
VERSION AM715484.1 GI:4037100
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 187)
Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
Two Neurospora crassa EST databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 152.
Location/Qualifiers
1. 187

BASE COUNT
69 a 35 c 23 g 60 t
ORIGIN

Query Match
89.3%; Score 13.4; DB 119; Length 151;

JOURNAL
COMMENT

Unpublished (1998)
Contact: Bruce A.

contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry

Department of Chemistry and Biochemistry
Advanced Center for Genome Technology

620 Parrington Oval, Norman, OK 73019, USA
Tel.: 405 325 4912

Tel: 405 325 4912
Fax: 405 325 7762

Fax: 405 325 7762
Email: bproe@ou.edu

Email: broe@ou.edu
We anticipate the

We anticipate the Genetics Stock Cen

Genetics Stock Center
Genetics Universal
Seq primer: Universal
Forward primer

seq primer: Universal
High quality sequence

high quality sequence stop: 273.
Location/Qualifiers

location/qualifiers
1. .323

FEATURES

source	location/qualifiers
1. .323	

/organism

Neurospora Class
strain-"bd, irq7 A"

/db_xref="laxon:5141"

```
clone="a2f03nm"
```

```
/clone_lib-"Neur
```

/tissue_type-"tissue harvested following 22hr growth in dark"

dark" \rightarrow "Vector" \rightarrow "growth" \rightarrow "dark"

```
/note="Vector: pBluescript SK-"; site_1: xbaI; site_2:
ECOR1. See: Ball-Podaras D.
```

5' end of cDNA cloned into pCR2.1 vector. See: Bell-Pedersen, D., et al. PNAS 93:13096, 1996

end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript.

end of cDNA cloned into EcoRI site of pBluescript"

Year	1990	2000	2010	2020
59 c	42 g	111 t		

BASE COUNT
ORIGIN

111 a	59 c	42 g	111 f
-------	------	------	-------

ORIGIN

Query Match

Query Match	89.3%;	Score 13.4;	DB 18;	length 323;
Best local Similarity	93.3%;	Pred. No. 4	50/02;	

	Similarity	95.38;	Pred. No.	4.5e102;	
Matches	14;	Conservative	0;	Mismatches	7;
				Indels	0;

QY 1 agcgggtctattaga 15

49 ACCORDANCE WITH THE

Db 49 ACCGGGCTCTATTATA 35

Search completed: May 1, 2001, 17:25:39
Job time: 8725 sec

Job time: 8725 sec